1981

The effects of varying amounts of visual information on telephone intelligibility

Debra N. Levine

Mary C. Saggau

"THE EFFECTS OF VARYING AMOUNTS OF VISUAL INFORMATION ON TELEPHONE INTELLIGIBILITY"

Debra N. Levine
Mary C. Saggau
Independent Study
May, 1981

Telephone communication is difficult for hearing-impaired people who cannot perceive speech reliably through listening alone. Many of these individuals, however, successfully communicate through speech in face to face situations and could benefit considerably from a video telephone channel. Previous work suggests that incomplete facial/oral images can be adequate complementary signals for audio-visual (A-V) communication by those who hear only low frequency cues (PPR #21, pp. 45-50; 687).

Work done at Bell Laboratories suggests that gross body movements may be depicted by trajectories of small lights or reflectors marking critical articulatory points (e.g., reflectors may be attached to finger joints to display elements of sign language). It is very likely that such reduced cues, coded in terms of X-Y coordinates, can be transmitted over a narrow bandwidth telephone line and later decoded and displayed as a facsimile of the original pattern.

In this study, we are determining whether similarly reduced amounts of visible articulatory information has potential for increasing speech intelligibility over the telephone to hearing-impaired listeners. A closed circuit television system (camera, TV monitor) and an audio channel (Bell Telephone Teletrainer) has been connected between two separate laboratories in the CID school. Pilot work has involved eight normal hearing

adults who listen to telephone output via an audio distortion system while visually attending to a TV monitor. Talkers have attempted to communicate syllables/words/sentences under different conditions at different times: (1) DP 30 Plexiglass placed before TV monitor screen; (2) TV camera lens defocused; (3) TV monitor contrast adjusted to minimum; (4) video signal low pass filtered; (5) white clown make up accentuating lips only; face and teeth blacked out; (6) fluorescent lipstick; ultraviolet light source; (7) various cardboard masks (apertures, slits, hole patterns) placed over TV monitor screen. Given a distorted audio signal that alone yielded near zero auditory recognition, A-V repetition of connected discourse ("tracking procedure", DeFilippo and Scott,1978) under all of these poor optical conditions was remarkably good, ranging from 30-50 words/minute (A-V normally is 70-80 words/minute with distorted audio and full video).

Considering the concurrent work at Bell Labs, we recently evaluated A-V intelligibility with a simple light pattern to indicate lip position and movement. After trial and error suggested optimal placement, six ¼" diameter discs punched from 3M reflective tape were affixed to the talker's lips as follows: 2 at midline (top, bottom), touching during bilabial closure; 2 each, top/bottom, 1 cm. to either side of midline. Light from a 9" diameter circular fluorescent tube was directed toward the mouth; the TV camera was aimed through the opening. When brightness and contrast controls on the TV monitor were suitably adjusted, only six small spots of light were visible.

The intent of the present study was to evaluate the ability of hearing-impaired listeners to understand speech over the telephone under three conditions: (1) acoustic signal alone; (2) acoustic signal plus full-face image of the talker; (3) acoustic signal plus image of six small moving lights reflected from the lips of the talker. The scores obtained under each of these three conditions were then compared to determine if the reduced visual information present in the third condition was sufficient for comprehension of speech over the telephone.

EXPERIMENT 1

The first section of this study involved two normally hearing subjects alternating as talker and listener. The acoustic signal received by the listener under each of the three conditions was first passed through a distortion system (see Figure 1). Each of the subjects served as listener for three 10 minute sessions under each of the three conditions. The sessions were limited to ten minutes each due to the high concentration level required when listening to distorted speech. Longer sessions may have caused a decrease in performance due to fatigue. The talker read to the listener from The Heart of the Matter by Graham Green (Viking Press, 1948). When the listener had difficulty in understanding part of the message, the "tracking" procedure (DeFilippo and Scott, 1978) was employed to achieve comprehension onthe part of the listener. Repetition or exaggeration of key words was the most commonly used strategy. A list of words starting with each of the consonants of

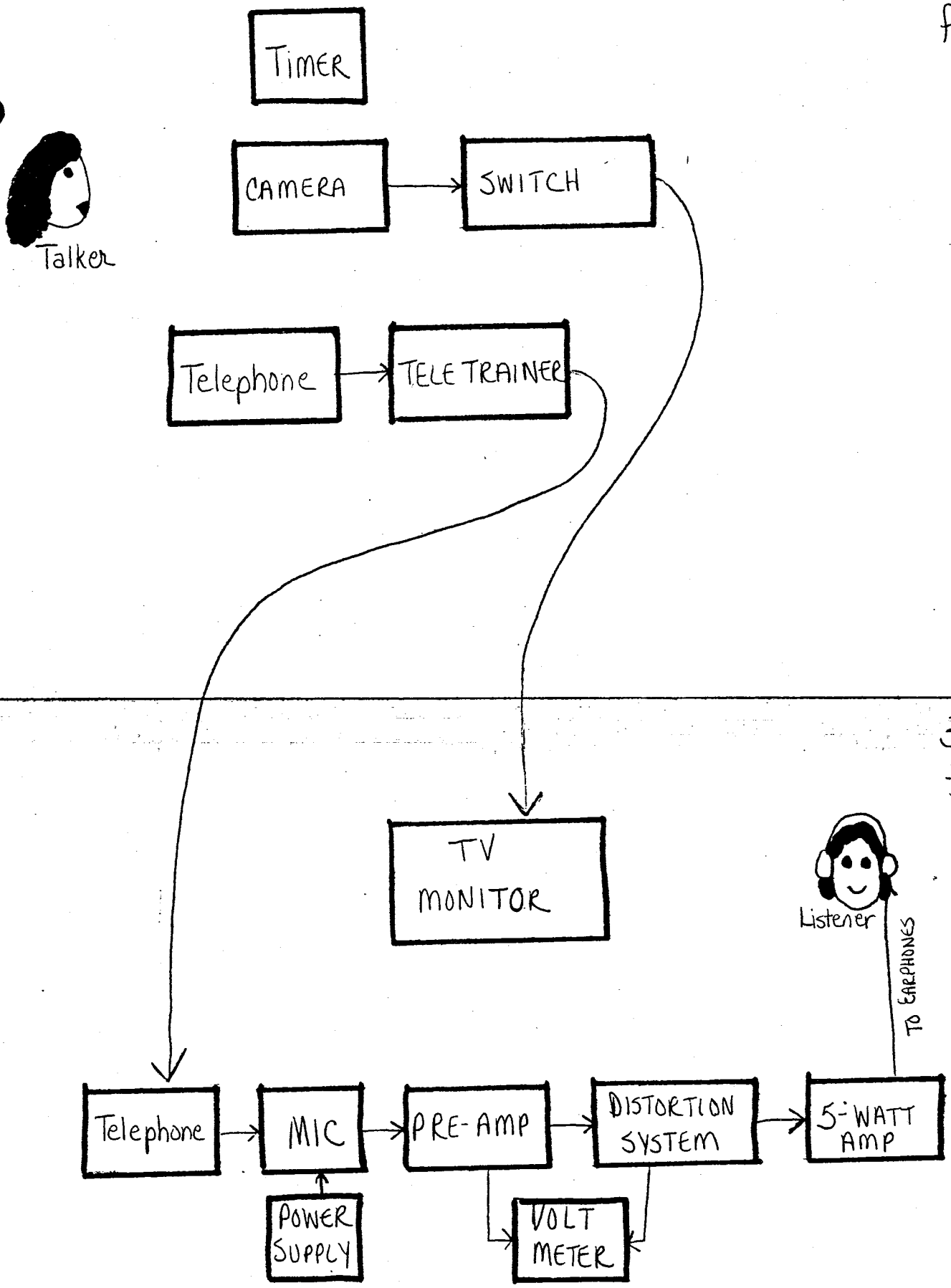


Fig. 1

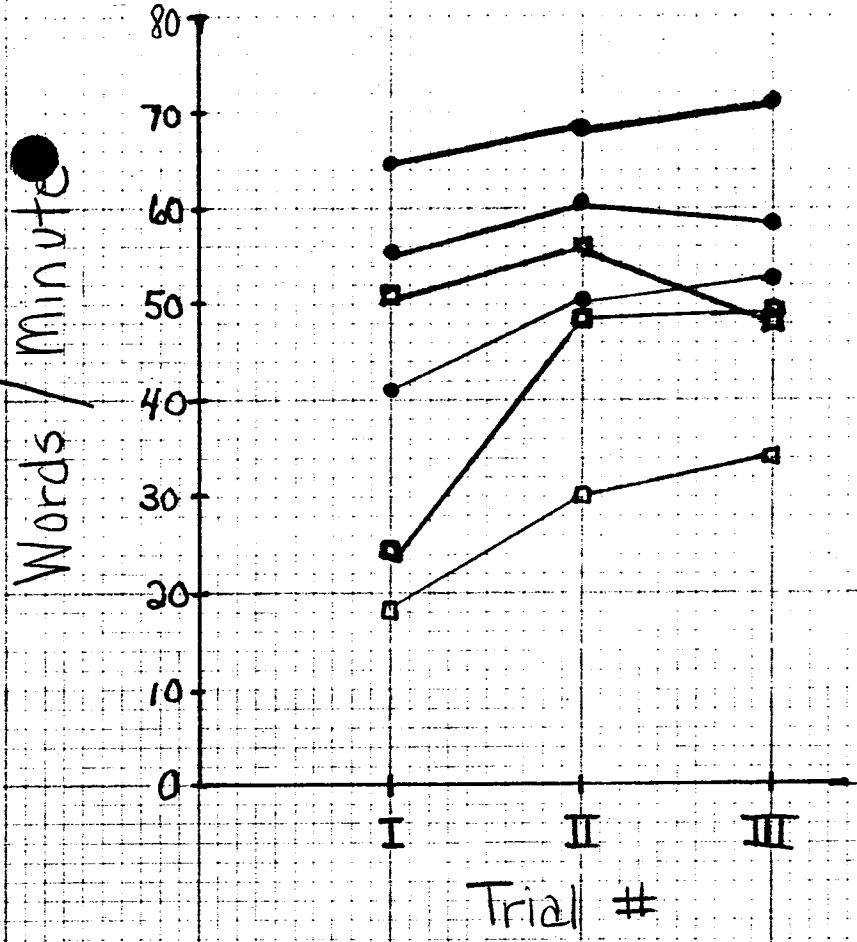
BLOCK DIAGRAM - 1ST SECTION OF EXPERIMENT

the alphabet, tested previously to insure intelligibility, was used when the listener had difficulty perceiving one particular sound (see Appendix). The strategy used for misperceived vowels was a coded number system (see Appendix). If the repetition strategy failed, the words were often spelled out using the consonant word list and the coded number system for vowels. It was found that using the key word in a sentence or phrase rarely increased the listener's comprehension. For example, if the misunderstood word was "bicycle", use of the key words in a phrase such as "bicycle built for two" did not help the listener. Even if the phrase was familiar to the listener under normal listening conditions, she frequently had no idea of the context and after it had passed through the distortion system, the phrase became practically unitelligible.

The scores were expressed in number of words repeated correctly per minute for each of the nine sessions per talker. A comparison of the scores obtained under each of the three conditions was made (see Figure 2).

When listening to the distorted speech, especially when receiving the acoustic signal alone, the listener had an extremely difficult time following the storyline. In most cases the listener was unable to relate what had happened in the section of the book that had been read during the session. However, when the subjects conversed over the telephone with the distortion system in a "normal" manner (listening alone), the conversation could be more easily followed.

At times, the listener repeated nonsense syllables which the



	Mary	Debbe
Auditory + Full Face	□	●
Auditory + 6-dots	□	●
Auditory	□	●

(Wds / Minute)

	Trial I		Trial II		Trial III	
	MARY	DEBBE	MARY	DEBBE	MARY	DEBBE
Auditory + Full Face	51	64.6	56.6	68.7	48	71.1
Auditory + 6 dots	24.4	55.7	48.9	60.5	49.2	58.4
Auditory	18.5	41.2	30.2	50.3	34.2	53.1

Fig. 2

talker indicated as being correct. These mistakes on the part of the talker were probably due to the distortion and static in the telephone receiver, causing the incorrect response of the listener to be perceived as correct by the talker.

As the listener gained experience in this task, scores improved. The scores received for the condition of acoustic signal plus the image of six small moving lights were much closer to those received for the condition of acoustic signal plus full-face image of the talker than to the scores received for the acoustic signal alone. It was anticipated that if the study included a greater number of sessions, the scores would improve even more.

While the distortion system used in this experiment may not exactly represent how speech sounds to the hearing impaired listener, the results indicate that even for inexperienced lipreaders listening to unfamiliar distorted speech, the visual speech information can be greatly reduced, i.e. to only 6 points of light, with only a moderate decrement in Audio-visual efficiency.

The purpose of experiment 1 was for the experimenters to gain experience with the listening conditions so that in the second part of the experiment, when school age children were used as subjects, they could be of assistance when difficulty in communication arose.

EXPERIMENT 2

The second section of this study involved ten 11 to 16 year old, severely hearing impaired CID students, who are prelingually deaf and proficient lipreaders. The subjects were chosen for this

experiment on the basis of their scores for a monosyllable-trochee-spondee recognition task. These subjects had average scores in order to insure that they could understand some speech acoustically, but would not understand everything that was said to them via the acoustic signal alone. The scores for the better ear of the subjects were no lower than 3 and no higher than 27.

The subjects listened over the telephone under the same conditions described previously. The arrangement of the Laboratory equipment is shown in Figure 3. The stimuli in the second section of the experiment, however, were 435 simple questions with one word answers. (Similar to PAL Auditory Test #12 - See Table 1) The majority of the questions were divided into semantic categories, i.e. body parts, colors, numbers, days, etc. Approximately the same number of questions from each of these categories was represented in each of the three listening conditions. However, since more questions were answered under the condition of acoustic signal plus full-face image of the talker than the other two conditions, a greater number of questions was required for this condition. The questions were selected and asked at random for a set, thus preventing the listener from guessing the answer from contextual cues.

The subjects were instructed to answer each question as quickly as possible. If they did not understand what was said, they were told to repeat the part of the question understood, if any, or to ask for a repetition of the entire question. A second experimenter sat in the room with the subject, listening to the talker over earphones. This experimenter kept track of the strategies used by the talker and the observer to achieve comprehension of the

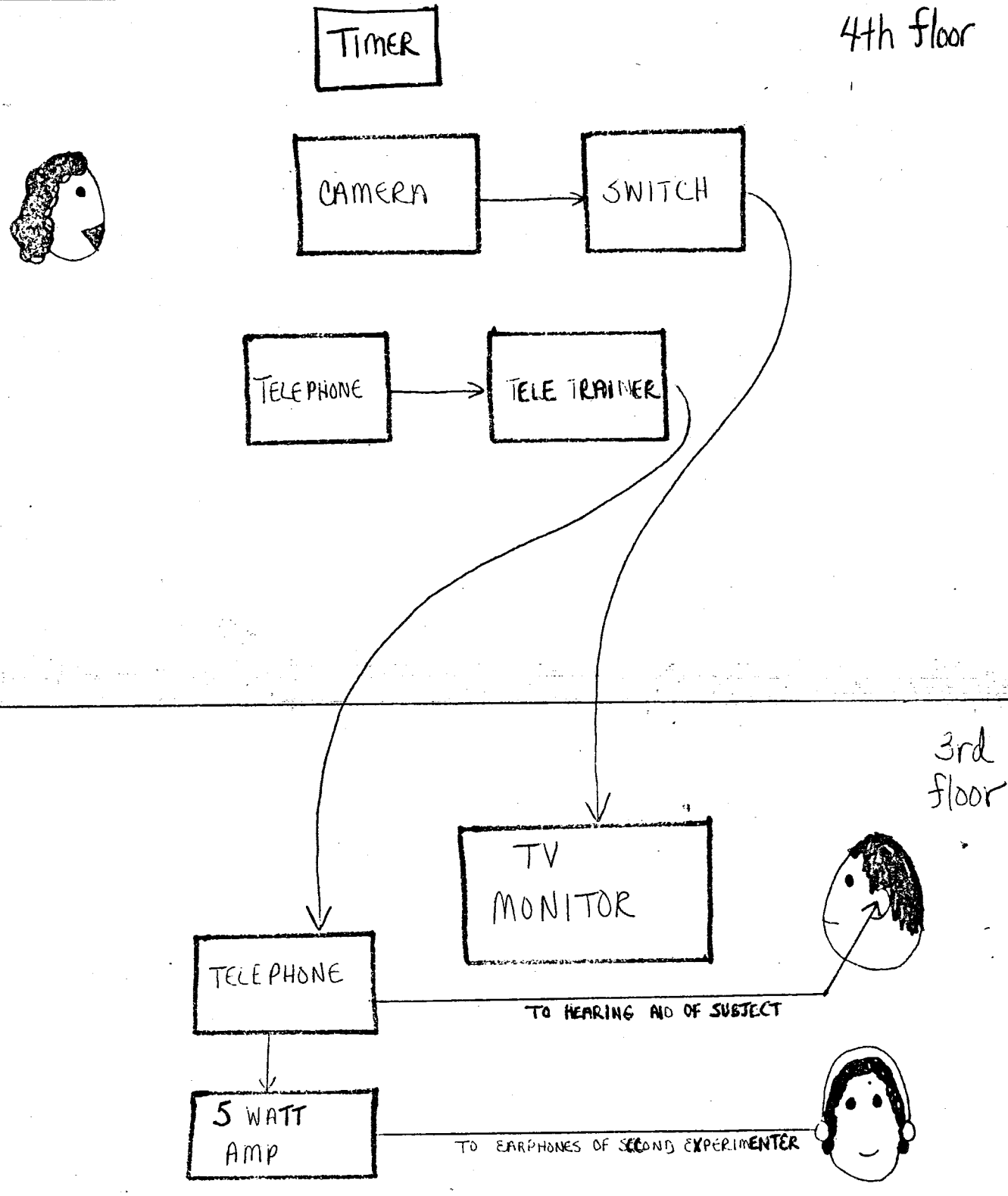


Fig. 3

BLOCK DIAGRAM - 2nd Section of Experiment

Table 1. Examples of stimulus questions used in the second section of the experiment.

What color is the sun?

How many legs do you have?

Who puts out fires?

What number comes after seven?

What do you use to sweep the floor?

What grows on top of your head?

What color is mustard?

What month comes between May and July?

What do you use to brush your teeth?

Is sugar sweet or sour?

question.

When the listener did not understand what was said, the assistant suggested various strategies. The strategy most commonly used by the talker was repetition of the question. If the listener understood only part of the question, the section that was not understood was repeated. If after two complete or partial repetitions, the listener still was unable to comprehend any part of the question, one of two strategies was employed: 1) Either the key words in the question were repeated separately (eg. What day comes after Monday? - day, after, Monday were repeated); or 2) The question was read word by word with the subject repeating what had been said after each word. The amount of time spent on each strategy varied from question to question, from condition to condition, and from subject to subject. If the experimenter in the room with the subject felt that the subject was on the verge of comprehension, the time allotted for the strategies was prolonged. Conversely, if it was felt that further attempts to help the subject understand would be futile, the talker was instructed to proceed to the next question. These subjective decisions were based on attentiveness of the subject, current frustration level of both the talker and the subject, and the subject's persistance in achieving comprehension.

RESULTS

The scores were expressed in percent of questions comprehended per minute (re total presented) and also in the number of questions answered correctly per minute. (See Figures 4-9) When the scores

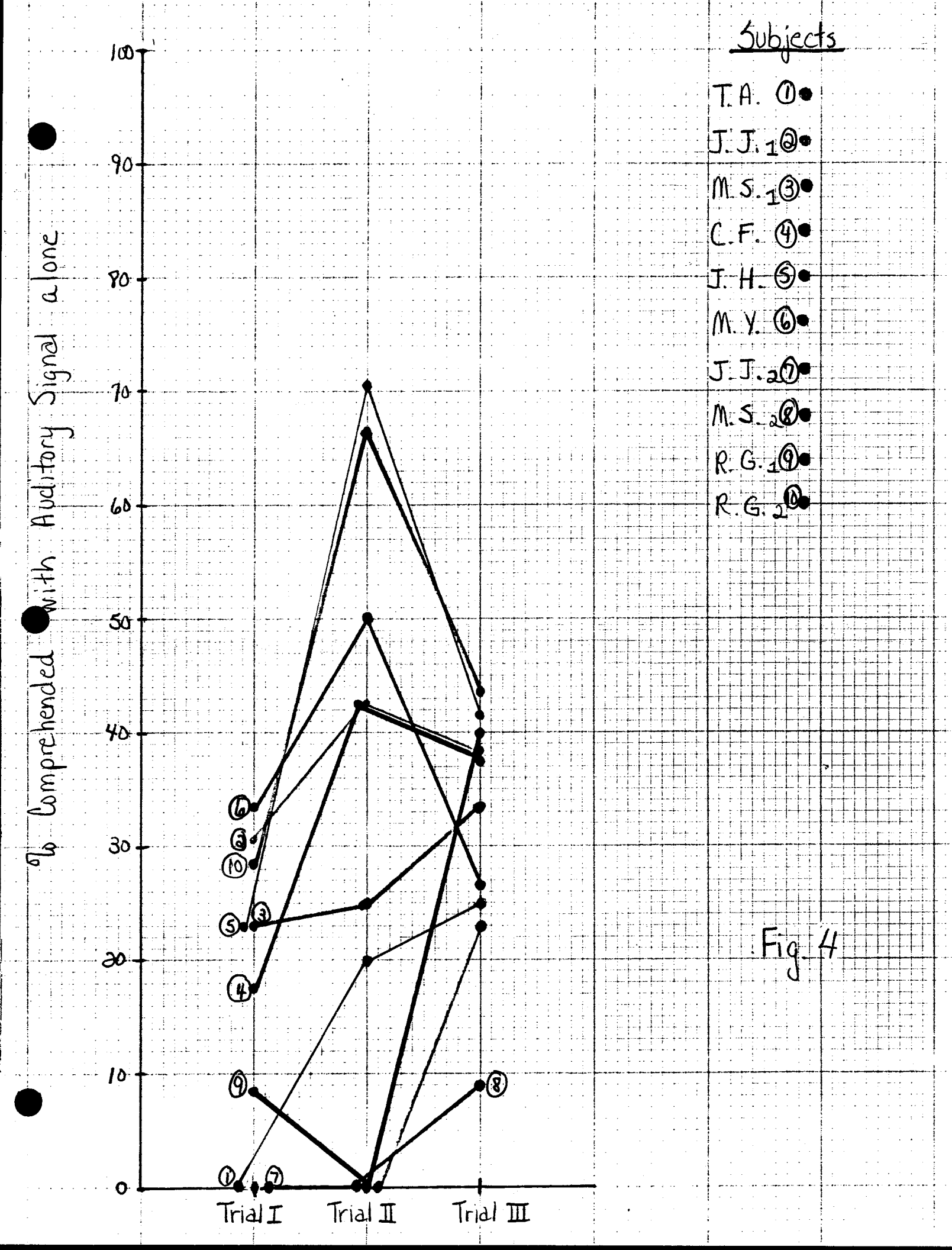

% Comprehended with Auditory Signal alone
100
90
80
70
60
50
40
30
20
10
0
Trial I
Trial II
Trial III
Subjects
T.A. ①
J.J.1 ②
M.S.1 ③
C.F. ④
J.H. ⑤
M.Y. ⑥
J.J.2 ⑦
M.S.2 ⑧
R.G.1 ⑨
R.G.2 ⑩

Fig. 4

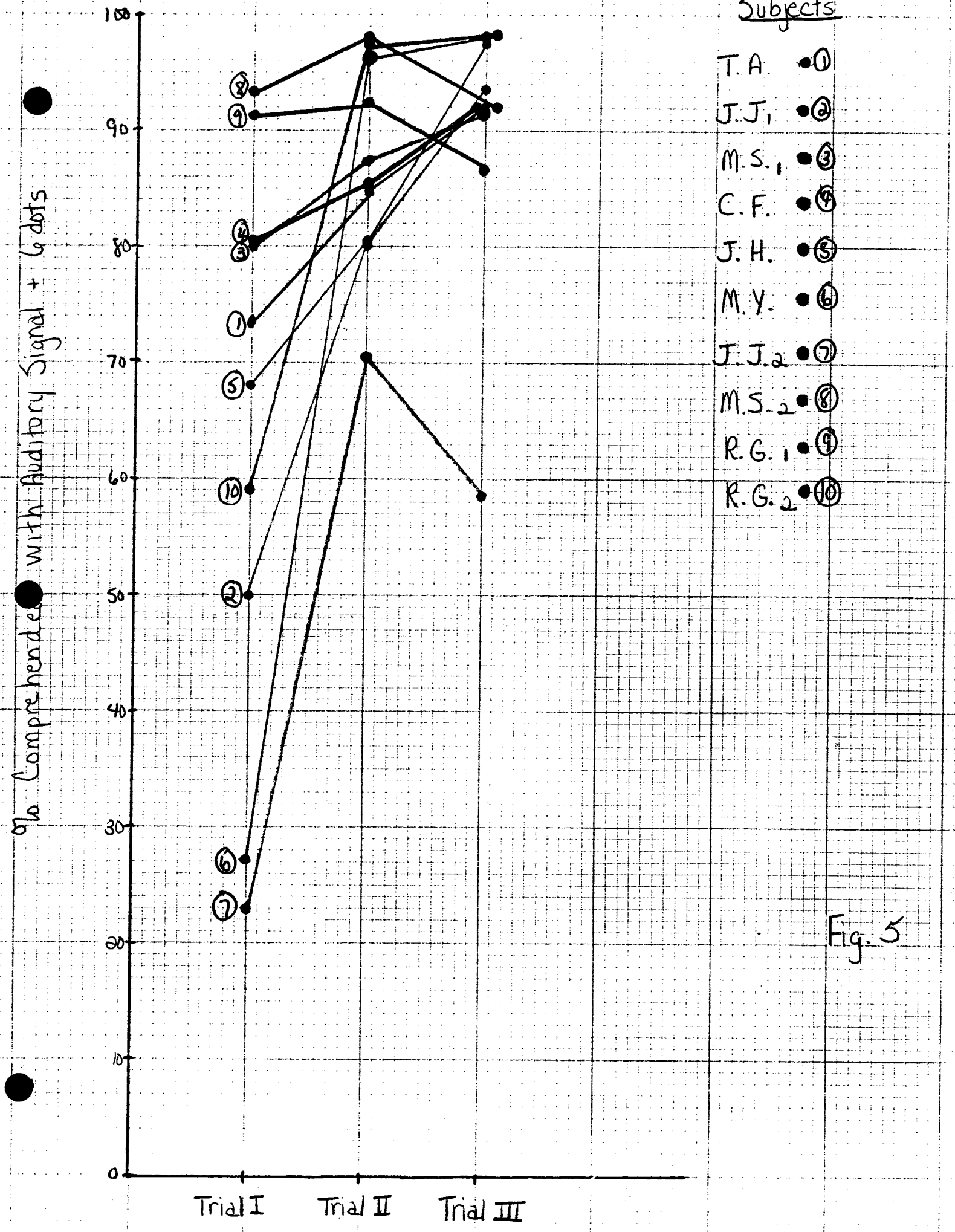

% Comprehended with Auditory Signal + 6 dots
100
90
80
70
60
50
40
30
20
10
0
Trial I
Trial II
Trial III
Subjects
T.A. ①
J.J.1 ②
M.S.1 ③
C.F. ④
J.H. ⑤
M.Y. ⑥
J.J.2 ⑦
M.S.2 ⑧
R.G.1 ⑨
R.G.2 ⑩

Fig. 5

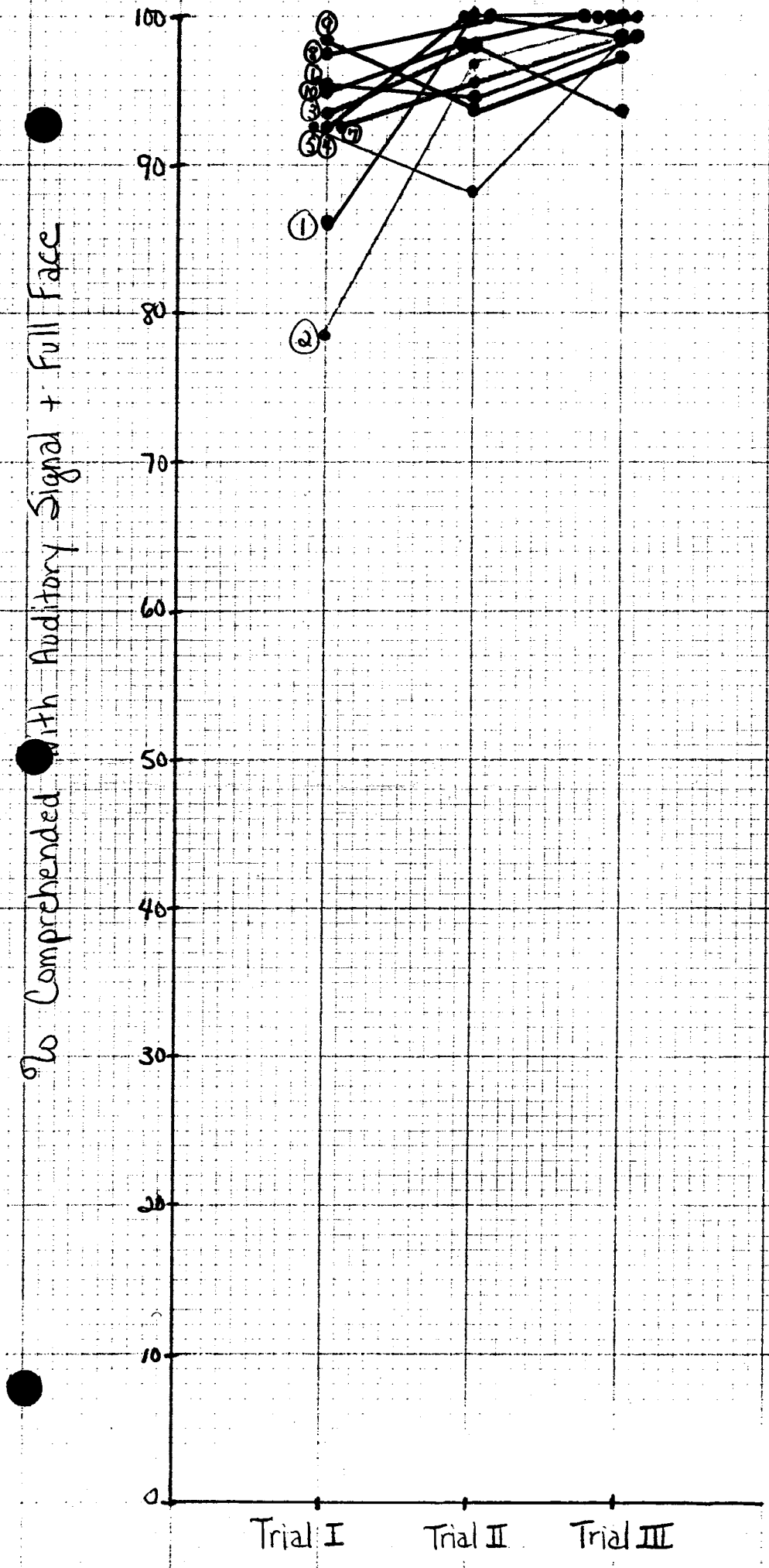

% Comprehended With Auditory Signal + Full Face
100
90
80
70
60
50
40
30
20
10
0
Trial I
Trial II
Trial III
Subjects
T.A. ①
J.J.1 ②
M.S.1 ③
C.F. ④
J.H. ⑤
M.Y. ⑥
J.J.2 ⑦
M.S.2 ⑧
R.G.1 ⑨
R.G.2 ⑩

Fig. 6

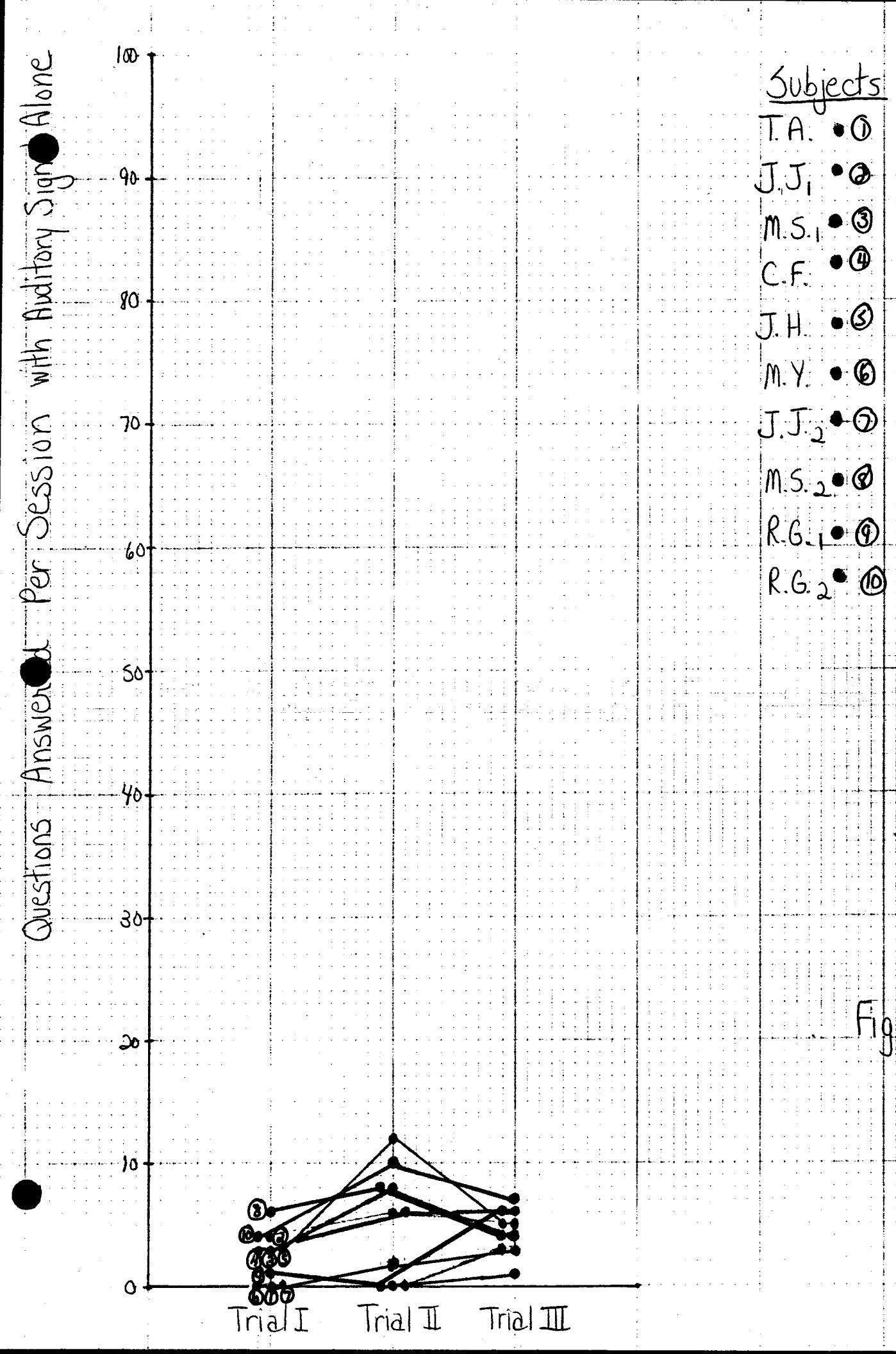

Questions Answered Per Session with Auditory Signal Alone
100
90
80
70
60
50
40
30
20
10
0
Trial I
Trial II
Trial III
Subjects
T.A. ①
J.J.1 ②
M.S.1 ③
C.F. ④
J.H. ⑤
M.Y. ⑥
J.J.2 ⑦
M.S.2 ⑧
R.G.1 ⑨
R.G.2 ⑩

Fig. 7

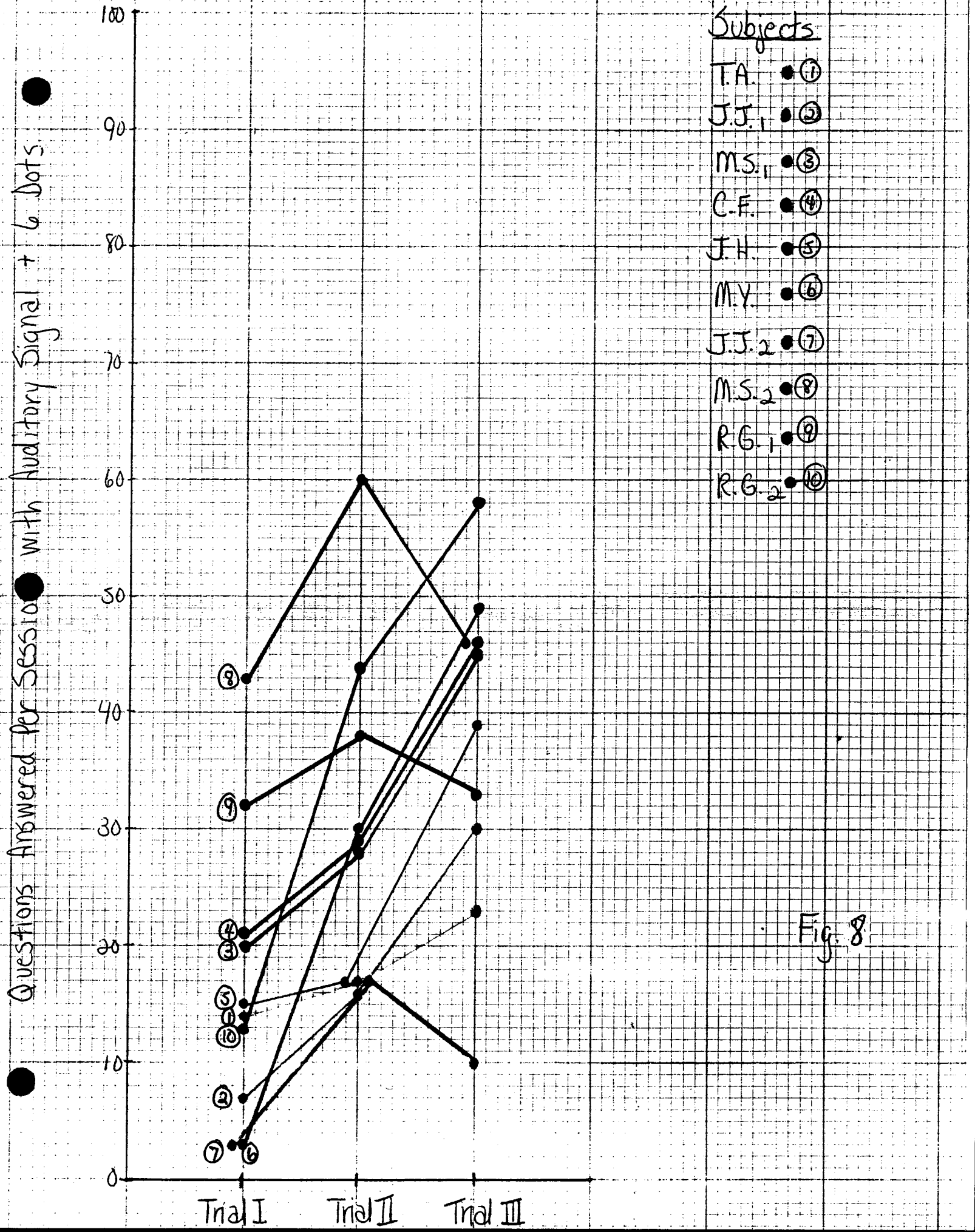

Questions Answered Per Session With Auditory Signal + 6 Dots
100
90
80
70
60
50
40
30
20
10
0
Trial I
Trial II
Trial III
Subjects
T.A. ①
J.J.1 ②
M.S.1 ③
C.F. ④
J.H. ⑤
M.Y. ⑥
J.J.2 ⑦
M.S.2 ⑧
R.G.1 ⑨
R.G.2 ⑩

Fig. 8

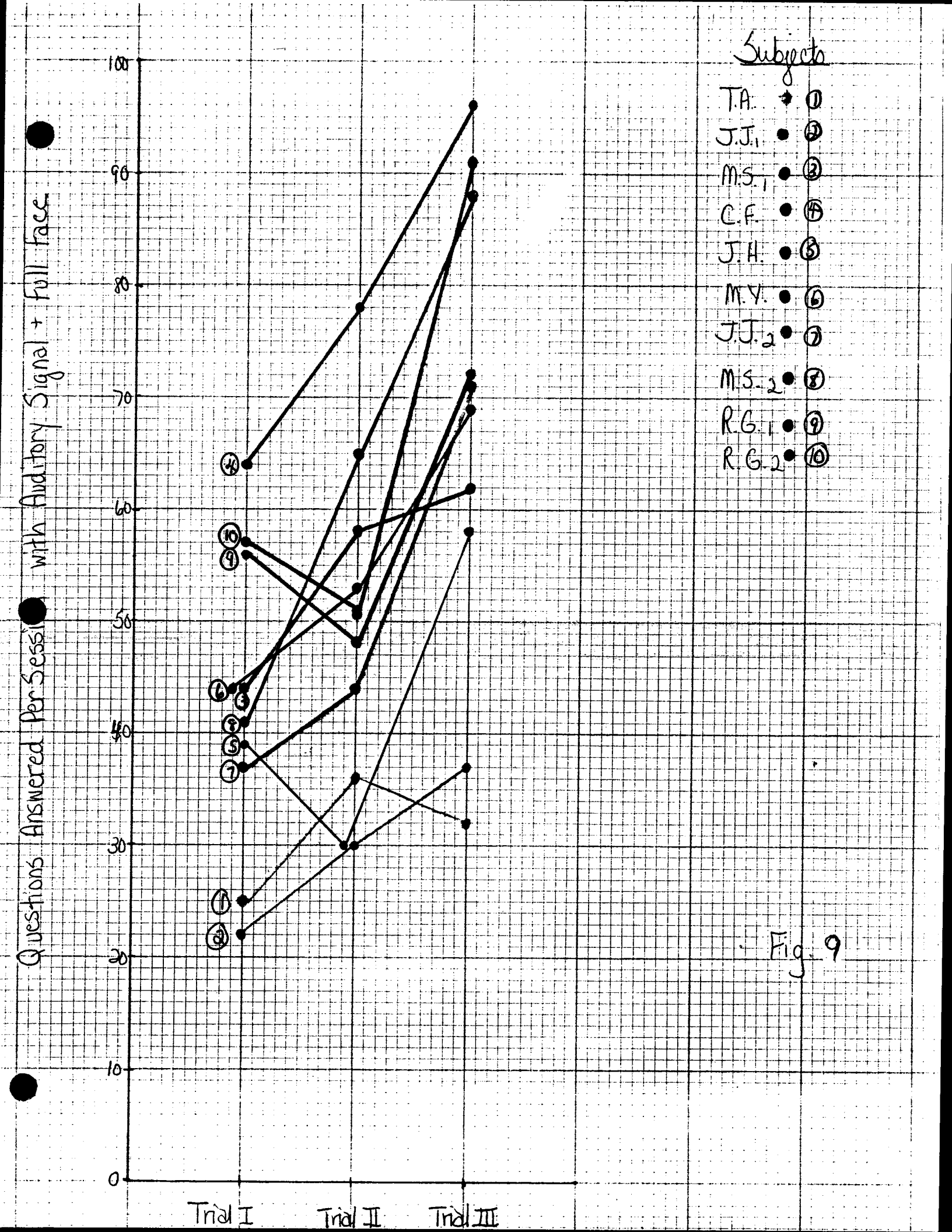

Questions Answered Per Session with Auditory Signal + Full Face
100
90
80
70
60
50
40
30
20
10
0
Trial I
Trial II
Trial III
Subjects
T.A. ①
J.J.1 ②
M.S.1 ③
C.F. ④
J.H. ⑤
M.V. ⑥
J.J.2 ⑦
M.S.2 ⑧
R.G.1 ⑨
R.G.2 ⑩

Fig. 9

are averaged across trials and across subjects the results show that when listening alone, the subjects understood only 27.2% of the questions asked. While listening and watching the full-face image of the speaker, the subjects could understand 95.6% of the questions asked. The average score for listening while watching the six moving spots of light was 78.4% which shows great improvement when compared to the condition of listening alone. These results indicate that a reduced amount of visual information, along with the auditory signal enabled these hearing impaired listeners to communicate over the telephone.

DISCUSSION

The ten students chosen for this study had no previous experience with listening over the telephone while attending to an optical image of six small moving lights. A large improvement in scores was noted between the first and second sessions and in most cases between the second and third session under this condition, which indicates learning. Three of the ten subjects exhibited a small decrease in scores between the second and third session. This reduction may be due to a decrease in enthusiasm on the part of the listener.

The subjects were not given specific training as to what to look for to identify various sounds (eg. the disappearance of the three lower dots when the consonants /f/ or/v/ were produced). Had they been instructed in this manner, scores may have been even greater for this particular condition. With additional practice, it is assumed scores would improve.

Since the questions asked were selected at random, the listener was prevented from guessing the answer from context. If it was unclear whether the listener understood the question or whether he was just guessing, he was requested to repeat what he had heard. Usually, in an average telephone conversation, the listener is provided with many contextual cues or the context is often provided by the listener himself, who asks clarifying questions in return. Therefore, it is expected that comprehension of a normal telephone conversation would be much greater for these subjects than what our results show for these isolated questions.

The results support the notion that severely hearing impaired people may be able to use the telephone with the addition of a limited amount of visual articulatory information (eg. six small moving lights.) It is hoped that these results would encourage development of a prototype system for transmitting optical components of speech over ordinary telephone lines.

Appendix

Consonant Word List

A-alphabet
B-bubble bath
C-Connecticut
D-Donald Duck
F-french fries
G-garbage can
H-hotdog
J-July
K-kangaroo
L-lollipop
M- marmalade
N-New York
P-Ponderosa
Q-Queen Elizabeth
R- raspberry
S-Sarasota
T-Tinkerbell
V-volcano
W-watermelon
X-eXcellent
Y-Yosemite Sam
Z-zebra

Coded Number System for Vowels

1. /i/ eat
2. /I/ hit
3. /e/ ape
4. /ɛ/ red
5. /æ/ hat
6. /a/ father
7. /ɔ/ bought
8. /o/ goat
9. /ʊ/ boot
10. /u/ book
11. /ɚ/ mother
12. /ʌ/ up

Dipthongs

1. /aʊ/ how
2. /ɔɪ/ boy
3. /eɪ/ ate
4. /aɪ/ kite
5. /ju/ you

BIBLIOGRAPHY

Central Institute for the Deaf Research Department Periodic Progress Report #21, pg. 45, 1977-1978.

Davis, H. and S.R. Silverman, Hearing and Deafness, Holt, Rinehart, and Winston, New York, 1978.

DeFilippo, C.L. and B.L. Scott, "A Method for Training and Evaluating the Reception of Ongoing Speech", Journal of the Acoustical Society of America, vol. 63, pgs. 1186-1192, 1978.